DOCUMENT RESUME

ED 038 269 SE 007 565

AUTHOR	Zunde, Pranas; Dexter, Margaret E.
TITLE	Indexing Consistency and Quality.
INSTITUTION	Georgia Inst. of Tech., Atlanta. School of Information Sciences.
SPONS AGENCY	National Science Foundation, Washington, D.C.
REPORT NO	RR-GITIS-69-08
PUB DATE	69
NOTE	42p.
AVAILABLE FROM	Clearinghouse for Federal Scientific and Technical Information, Springfield, Va. 22151
EDRS PRICE	EDRS Price MF-$0.25 HC-$2.20
DESCRIPTORS	Classification, *Indexing, *Information Retrieval, *Information Science, *Subject Index Terms

ABSTRACT

Proposed is a measure of indexing consistency based on the concept of "fuzzy sets." By this procedure a higher consistency value is assigned if indexers agree on the more important terms than if they agree on less important terms. Measures of the quality of an indexer's work and exhaustivity of indexing are also proposed. Experimental data on indexing consistency is presented for certain categories of indexers, and consistency, quality, and exhaustivity values are compared and analyzed. The analysis of indexing exhaustivity leads to the conclusion that the increase of information as a result of group indexing obeys the same law as that of the scattering of information (Bradford), of scientific productivity (Lotka), and vocabulary distribution (Zipf). (RP)

Work reported in this publication was performed at the School of Information Science, Georgia Institute of Technology. The primary support of this work came from the Georgia Institute of Technology and from agencies acknowledged in the report.

* * *

The graduate School of Information Science of the Georgia Institute of Technology offers comprehensive programs of education, research and service in the information, computer and systems sciences. As part of its research activities the School operates, under a grant from the National Science Foundation, an interdisciplinary science information research center. Correspondence concerning the programs and activities of the School may be addressed to Director. School of Information Science, Georgia Institute of Technology, Atlanta, Georgia 30332.

Telephone: (404) 873-4211

GITIS-69-08

(Research Report)

INDEXING CONSISTENCY AND QUALITY

Pranas Zunde

Margaret E. Dexter

School of Information Science 1969

GEORGIA INSTITUTE OF TECHNOLOGY

Atlanta, Georgia

ACKNOWLEDGEMENT

The work reported in the paper has been sponsored in part by the National Science Foundation Grant GN-655. This assistance is gratefully acknowledged.

Vladimir Slamecka

VLADIMIR SLAMECKA
Project Director

Abstract

A measure of indexing consistency is developed based on the concept of "fuzzy sets". It assigns a higher consistency value if indexers agree on the more important terms than if they agree on less important terms. Measures of the quality of an indexer's work and exhaustivity of indexing are also proposed. Experimental data on indexing consistency is presented for certain categories of indexers, and consistency, quality, and exhaustivity values are compared and analyzed. The analysis of indexing exhaustivity leads to the conclusion that the increase of information as a result of group indexing is a process analogous to the Bradford's law of information scattering, Lotka's law of scientific productivity, and Zipf's law of vocabulary distribution.

INDEXING CONSISTENCY AND QUALITY

by

Pranas Zunde and Margaret E. Dexter*)

Definition of Indexing Consistency. It is well known that any two indexers, indexing one and the same document individually, will select sets of indexing terms which are most unlikely to be identical. In other words, if one compares indexing terms assigned by any two indexers to the same document, one discovers that, as a rule, the indexers differ considerably in their judgment as to which terms reflect the contents of the document most adequately. It is clear that this difference in the judgment of indexers introduces a great deal of uncertainty in any information retrieval system based on human indexing. Although an information system normally would not contain identical documents indexed by different indexers, the implication is that documents which are similar may and often would be indexed so differently that their similarity would not be properly reflected in the sets of indexing terms assigned to these documents. Various tools have been developed to reduce this discrepancy in the judgment of, i.e. term assignments by, the indexers, but the discussion of these tools is not the purpose of this article.

Since indexing consistency manifests itself in the similarity (or dissimilarity) of indexing terms assigned to a given document by different indexers, and since the selection of indexing terms by an indexer reflects his judgment regarding the information contained in the document and its representation, indexing consistency is essentially a

---

*)Georgia Institute of Technology, School of Information Science.
This research was partially supported by the NSF Grant GN-655.

measure of the similarity of reaction of different human beings processing the same information. Thus, more precisely, we shall define indexing consistency in a group of indexers as the degree of agreement in the representation of the essential information content of the document by certain sets of indexing terms selected individually and independently by each of the indexers in the group.

Previous Consistency Studies. In a survey of indexing consistency studies, Hooper (1) cited consistency values ranging from 10% to 80%, depending on the conditions under which the indexing was performed and the measure of consistency used. Jacoby (2) reported a mean consistency of 10% when chemical patents were indexed by three experienced and three inexperienced indexers, and no indexing aids were used. Consistency values of 35% to 45% were obtained by Slamecka and Jacoby (3) for experienced indexers using certain indexing aids, such as controlled vocabulary.

In a third report, Jacoby and Slamecka (4) arrived at a consistency of 16.3% for experienced indexers and 12.6% for inexperienced indexers. No indexing aids other than indexing rules were used.

Painter (5) reported consistency values of 40%, 42%, 48% and 70% with varying indexing systems and types of documents. Rodgers (6) came up with an average consistency of 24% for combinations of two indexers in a group. Korotkin and Oliver (7) reported consistencies ranging from 36% to 59% in an experiment in which five psychologists and five non-psychologists indexed abstracts.

Consistencies up to 80% were reached when indexers were required to use classification schedules and thesauri and had very limited freedom or no freedom at all to use terms not contained in the above indexing aids (5, 7, 8, 9, 10).

Schultz, Schultz and Orr (11) measured the goodness of author indexing by using a set of indexers as a criterian group to establish a weight for each term. The weight for a term for a given document was defined as the square of the number of criterian group indexers selecting the term for the specific document. The goodness of the author's index set was taken to be the sum of the weights of the terms selected. This measure was then normalized by dividing the score for each document by the maximum score possible for the document, the sum of the weights of all terms selected by the criterian group for the document. They found that the mean score for the authors was 76%.

Consistency Measure Used in Previous Studies. The consistency measure used in most previous studies was the ratio of the number of terms selected by all indexers in the group to the total number of different terms selected for the document. For a group of two indexers the consistency measure on a given document is thus defined as

$$c_{ij} = \frac{n(T_1 \cap T_2)}{n(T_1 \cup T_2)} \qquad (1)$$

where T_1 and T_2 denote the sets of terms selected by the first and second indexers respectively, and $n(T)$ denotes the number of elements in set T.

This measure can be extended to $k > 2$ indexers by taking

$$c_{1,2,\ldots,k} = \frac{n(T_1 \cap T_2 \cap \ldots \cap T_k)}{n(T_1 \cup T_2 \cup \ldots \cup T_k)} \qquad (2)$$

The main disadvantage of such a consistency measure is the underlying assumption that all the indexing terms selected are equally significant and relevant for the representation of the information content of the document. In other words, measures Eq(1) and Eq(2) completely disregard the difference between agreement on significant and the agreement on insignificant terms. The result, as will be shown in this paper, is that the above measures tend to produce consistently lower consistency measures than one would intuitively be willing to accept. For example, consider three indexers indexing (one and the same) document on jet propulsion and assume, for simplicity, that each of them assigned three indexing terms to the document as shown below

Indexing Term	Indexer		
	1	2	3
jet	x	x	x
propulsion	x	x	
flow	x		x
level		x	x

Assume further that the terms JET and PROPULSION refer to the main topic of the document, whereas FLOW and LEVEL refer to topics of marginal importance. From Eq(1), the consistency of any two of these three indexers is 0.50 (or 50 percent). But in effect the indexing consistency of 1 and 2 should be considered much higher than either that of 1 and 3 or of 2 and 3 since the common part of sets I_1 and I_2 above contain both of the significant words JET and PROPULSION. Therefore the measures Eq(1) or (2) would be fair measures of consistency only in the absence of any information whatsoever on the relative importance of the indexing terms, but they obviously do not adequately reflect the agreement of indexers' judgment if such information is available.

Proposed Consistency Measure. The consistency measure which is proposed here and which is expected to eliminate, at least partially, the shortcomings of the consistency measures of the type expressed by Eq (1) and Eq (2), is based on the postulate that there exists no well defined set of "relevant", "most indicative", "most pertinent", "most informative", etc. indexing terms for a document, because there exist no objective criteria which would enable us to construct such sets. Indexing performance of human indexers demonstrates this clearly, because if such criteria were available, we could apply them to obtain 100 percent indexing consistency.

An alternative to a well defined set is a "fuzzy set" which has been proposed by Zadeh (12). Whereas well defined sets have precisely stated criteria of membership, a fuzzy set is a collection of objects which meet the criteria of membership to a varying degree which is assumed as given. More precisely, a fuzzy set A in a set X of objects x is characterized by a membership function f(x) which associates with each point in X a real number in the interval [0,1], with the value of $f_A(x)$ at x representing the "grade of membership" of x in A. The nearer the value of $f_A(x)$ to unity, the higher the grade of membership of x in A. If the function $f_A(x)$ can assume only the values 0 and 1, it reduces to the familiar characteristic function of the set A. For further discussion of the fuzzy sets and operations on them, see Appendix.

Consider now the set of all English words and phrases, which are potential indexing terms for any kind of document. We shall say that this set of words and phrases is a fuzzy set with respect to the membership criteria of "being representative of or pertinent to" a given document D , the degree of membership of each term reflecting the degree of agreement as to its significance with respect to the information it conveys about the document indexed. We shall call this set the global (indexing) set U for the document D .

We shall further define the set T_j of indexing terms t assigned by an indexer I_j to the document D to be the subset of the fuzzy set U such that for each $t \in T_j$, $f_{T_j}(t) = f_U(t)$. In other words, the set T_j is a fuzzy set obtained by associating with each indexing term t selected by the indexer I_j the membership value which that term t has in the set U. For example, the set $T_1 = \{t_1, t_2, t_3, t_4\}$ = {blood, circulation, heart, disease} with the membership function $f(t_1) = 0.8$, $f(t_2) = 0.85$, $f(t_3) = 0.95$, $f(t_4) = 0.7$ might represent the set of indexing terms assigned by an indexer I_1 to some document D .

Now let $\{I_1, I_2, \cdots, I_m\}$ be a group of indexers and $\{T_1, T_2, \cdots, T_m\}$ a (well defined) collection of fuzzy sets of indexing terms t assigned by each of these indexers to one and the same document D. We shall define the measure of consistency of the group of indexers $I_1, I_2, \cdots, I_m$ by the expression

$$C^*_{1,2,\cdots,m} = \frac{\sum_t f_{T_1 \cap T_2 \cap \cdots \cap T_m}(t)}{\sum_t f_{T_1 \cup T_2 \cup \cdots \cup T_m}(t)} = \frac{\sum_t f_{\bigcap_m T_j}(t)}{\sum_t f_{\bigcup_m T_j}(t)} \qquad (3)$$

where $\sum_t f_{T_1 \cap T_2 \cap \cdots \cap T_m}(t) = \sum_t f_{\bigcap_m T_j}(t)$ denotes the sum of the membership values of the intersection set of the fuzzy sets T_j, $j = 1, 2, \cdots, m$, and $\sum_t f_{T_1 \cup T_2 \cup \cdots \cup T_m}(t) = \sum_t f_{\bigcup_m T_j}(t)$ denotes the sum of the membership values of the union set of the fuzzy sets T_j, $j = 1, 2, \cdots, m$.

Quality and Exhaustivity Measures. Furthermore we shall define the measure of quality of an index to some document D to be the expression

$$n_j = \frac{\sum_t f_{T_j}(t)}{\sum_t f_U(t)} \quad (4)$$

where $\sum_t f_U(t)$ denotes the sum of the membership values of all the elements t of the global set U,

and $\sum_t f_{T_j}(t)$ denotes the sum of the membership values of all the elements t of the fuzzy set of indexing terms T_j, assigned by the indexer I_j to the document D.

It should be noted that the measure of quality given by expression (3) can be viewed as the measure of goodness of performance of indexer I_j in indexing document D, where by performance is understood the capabilities of the indexer to select indexing terms which would give maximum information about the document indexed.

Finally we shall define the measure of exhaustivity by the expression

$$\gamma_{1,2,\ldots,m} = \frac{\sum_t f^*_{\bigcup_m T_j}(t) \ln f^*_{\bigcup_m T_j}(t)}{\sum_t f^*_U(t) \ln f^*_U(t)} \quad (5)$$

where

$$f^*_{\bigcup_m T_j}(t) = \frac{f_{\bigcup_m T_j}(t)}{\sum_t f_{\bigcup_m T_j}(t)} \quad (6)$$

and

$$f_u^*(t) = \frac{f_u(t)}{\sum_t f_u(t)} \tag{7}$$

Essentially this measure indicates that portion of the overall information of the document D which is reflected in the joint output of a group of indexers $\{I_1, I_2, \ldots, I_m\}$.

Since $\gamma_{1,2,\ldots,m} \geq \gamma_{1,2,\ldots,s}$ whenever $m > s$, i.e. γ increases or remains unchanged when the number of indexers in the group increases, the change in γ can be considered as the indicator of the proportion of additional information about the document gained due to the contribution of m-s additional indexers.

A problem of practical importance is how to obtain the global set U with its membership function for a document D . Leaving the question open, whether or not it is possible to determine such a set a priori, we shall describe an experimental approach to the construction of such a global set and development of the measures defined by Eq (3), Eq (4), and Eq (5), which enabled us to demonstrate the major issues of this investigation. The method used for the construction of the global set — and it is not claimed that this is the only method to arrive at such a set — will be outlined following the description of the experiments performed and data used.

Experimental Data. The data used in this investigation came from two sources. One source of data was a study performed by Schultz, Schultz and Orr (11) in which 285 biomedical documents were indexed by the author by twelve biomedical scientists who were engaged in research in the area, and by eight professional indexers. For the purposes of this study, data from twenty-nine of these documents was used. Two groups of eight indexers each were selected, the eight professional indexers forming one group and the first eight of the twelve scientists forming the other.

The documents indexed in this study were brief summaries of oral reports on current biomedical research which were presented at the 1962 meeting of the Federation of American Societies for Experimental Biology. These documents consisted of every tenth document in the 1962 meeting issue of the Federation Proceedings (Vol. 21, No. 2, March-April 1962.) The twelve scientists who indexed the documents were given a form which listed 373 subject categories from which they could select as many terms as they wished. The form also provided space for writing in additional terms. The scientists were instructed to choose as many terms as they felt were required to characterize the document. They were also told to index personally, to use the form only when it was natural and helpful and to let their responses reflect their viewpoints and terminology.

The eight professional indexers indexed the same documents using the same form. They were also encouraged to assign to each document as many terms as they considered necessary.

The other source of data was an experiment performed at Georgia Tech in May of 1968. In this experiment nine graduate students in the School of Information Science each indexed sixteen documents. Eight of these documents were selected from the 285 biomedical documents described above. The other eight were selected from Efficient Reading, by James E. Brown (13), a collection of articles designed for use in a rapid reading course. Data from eight of the nine graduate students was used for this study. No classification schedules or any other indexing aids were made available to these students. No restrictions were imposed as to the number of terms they were to assign to the documents.

Thus in the first instance we had slightly controlled indexing, in the second, completely free indexing.

A sample document of the biomedical category is shown in Figure 1. Indexing terms which have been assigned to this document by student indexers are given in Table 1.

Figure 1. A sample document from the biomedical collection used in indexing experiment.

PATHWAYS OF INTRACELLULAR HYDROGEN TRANSPORT.
Bertram Sacktor, Arthur R. Dick* and Eva H. Wormeer*
CRDL, Army Chemical Center, Maryland.

The mechanisms for oxidizing extramitochondrial DPNH in mammalian tissues were studied. Activity measurements were made of the pathways in different tissues, including: heart, skeletal muscle, diaphragm, brain, lung, liver, kidney, spleen and testes, for oxidizing exogenous DPNH by mitochondria and by cytoplasmic or soluble lactic, a glycerophosphate and malic dehydrogenases. Analyses of the oxidized and reduced metabolites in the different tissues were correlated with the potential of the different pathways.

Table 1. Terms Assigned by Student Indexers to Sample Document Shown in Figure 1.

Indexing term		Student indexer							
		1	2	3	4	5	6	7	8
t_1	analyses					x			
t_2	biochemistry		x				x		
t_3	cellular	x							
t_4	DPNH	x	x					x	x
t_5	extramitochondrial		x	x					
t_6	hydrogen				x	x	x		x
t_7	intracellular					x	x		
t_8	mammalian		x	x		x			
t_9	metabolic	x	x		x	x			x
t_{10}	mitochondria		x						
t_{11}	oxidase	x	x	x		x		x	x
t_{12}	pathways		x				x		
t_{13}	tissue		x		x	x			
t_{14}	transport				x		x	x	x

Derivation of Fuzzy Sets. For each document D , the global fuzzy set U was obtained by the following procedure. Every term assigned by any one of the indexers I_j, j = 1, 2,..., m, to the document D is an element of the fuzzy set U . The membership function of the global set U is obtained by assigning to each element, i.e. each term in the set, a membership value equal to the ratio of the number of indexers who assigned that term to the document D to the total number of indexers who indexed that document, i.e. total number of indexers in the experimental group.

The conceptual justification for this procedure is that the selection of a term by an indexer is an indication that that term is representative, at least in his judgment and to some degree of the information contained in the document and that it reflects some of that information. Hence every term selected by an indexer in the group should be assigned a membership value greater than zero. The rest of the terms in the vocabulary, i.e. the terms which none of the indexers used, are assumed to have the membership value equal to zero. As to the terms which have been selected, the degree of concensus of indexers in selecting a term is considered to be a proper indicator of their significance in representing the information contained in the document. In other words, the more indexers select a given indexing term, the more representative it should be considered with respect to the contents of the document. A term, which is selected by all indexers in the group, is assigned member-value 1. Thus, the membership values of the elements (terms) of the global fuzzy set U can assume values in the interval [0, 1], but terms with membership values equal to zero are not shown, since they have no effect on further calculations.

It should be noted that the fuzzy set U thus arrived at as well as the fuzzy sets T_j of indexing terms selected by each indexer in the experimental group are "fuzzy" with respect to their pertinence or significance in representing information contained in a document, but they are not "fuzzy" with respect to the criterion of being symbols for certain words, i.e. with respect to the partitioning of the vocabulary into a set of terms which have been chosen as indexing terms and a set of terms which have not been chosen as indexing terms. This is true for any fuzzy set: the concept of "fuzziness" pertains to the criteria of membership in a given set, and the same set could be a fuzzy set with respect to one criterion and a well defined set with respect to another.

As an example, the fuzzy global set U and three of the sets T_j the document D in Figure 1, calculated from the data in Table 1, are given below.

$$U = \left\{ \begin{array}{cccccccccccccc} t_1, & t_2, & t_3, & t_4, & t_5, & t_6, & t_7, & t_8, & t_9, & t_{10}, & t_{11}, & t_{12}, & t_{13}, & t_{14} \\ 1/8 & 2/8 & 1/8 & 4/8 & 2/8 & 4/8 & 2/8 & 3/8 & 5/8 & 1/8 & 6/8 & 2/8 & 3/8 & 4/8 \end{array} \right\}$$

$$T_1 = \left\{ \begin{array}{cccccccccccccc} t_1, & t_2, & t_3, & t_4, & t_5, & t_6, & t_7, & t_8, & t_9, & t_{10}, & t_{11}, & t_{12}, & t_{13}, & t_{14} \\ 0 & 0 & 1/8 & 4/8 & 0 & 0 & 0 & 0 & 5/8 & 0 & 6/8 & 0 & 0 & 0 \end{array} \right\}$$

$$T_2 = \left\{ \begin{array}{cccccccccccccc} t_1, & t_2, & t_3, & t_4, & t_5, & t_6, & t_7, & t_8, & t_9, & t_{10}, & t_{11}, & t_{12}, & t_{13}, & t_{14} \\ 0 & 2/8 & & 4/8 & 2/8 & 0 & 0 & 3/8 & 5/8 & 1/8 & 6/8 & 2/8 & 3/8 & 0 \end{array} \right\}$$

$$T_3 = \left\{ \begin{array}{cccccccccccccc} t_1, & t_2, & t_3, & t_4, & t_5, & t_6, & t_7, & t_8, & t_9, & t_{10}, & t_{11}, & t_{12}, & t_{13}, & t_{14} \\ 0 & 0 & 0 & 0 & 2/8 & 0 & 0 & 3/8 & 0 & 0 & 6/8 & 0 & 0 & 0 \end{array} \right\}$$

Results of Consistency Tests. Under all experimental conditions, the proposed measure of consistency, which reflects the agreement of a group of indexers on the significance of the selected terms, produced on the average higher consistency values than the measure Eq (2) which does not reflect any judgment of significance of the terms.

The consistency of one pair of professional indexers was calculated for the twenty-nine biomedical documents. The resulting mean and variance for both the "unweighted" measure c_{ij} and the proposed "weighted" measure c^*_{ij} are given in Table 2.

Table 2. Mean and Variance of Consistency of One Pair of Indexers over a Sample of 29 Documents for Consistency Measures c_{ij} and c^*_{ij}

	$c_{1\,2} = \dfrac{n(T_1 \cap T_2)}{n(T_1 \cup T_2)}$	$c^*_{1\,2} = \dfrac{\sum_t f_{T_1 \cap T_2}}{\sum f_{T_1 \cup T_2}}$
Mean	.24	.41
Variance	.01	.02

Consistency was then calculated for all pairs of indexers for a specific document. The average consistency according to the measure c_{ij} was 0.35, the average according to the proposed measure c^*_{ij} was 0.59. When random samples of pairs of indexes over all documents were calculated, the two consistency values were found to be 0.27 and 0.44 respectively.

When both the "unweighted" and the proposed "weighted" values were calculated for all pairs of professional indexers, it was found that the measure c_{ij} was higher in 2% of the cases, both measures were zero in 7½% of the cases, and the measure c^*_{ij} was higher for the remaining 89½% of the pairs of indexers.

The calculations were then extended to measure the consistency of three indexers, four indexers, up to the consistency of the entire group f eight indexers using Eq (2) and Eq (3), respectively.

Random combinations of three indexers, four indexers, and up to eight indexers were selected from one document. Again, the proposed measure c^*_{ij} tended to yield higher values of consistency than did the measure c_{ij}. The results are shown in Figure 2. Then the same calculations were made for combinations of indexers selected randomly from all documents. The results were quite similar to those for one document as may be seen in Figure 3.

Figure 2. Consistency of Indexing: Values averaged over combinations of indexers for one document

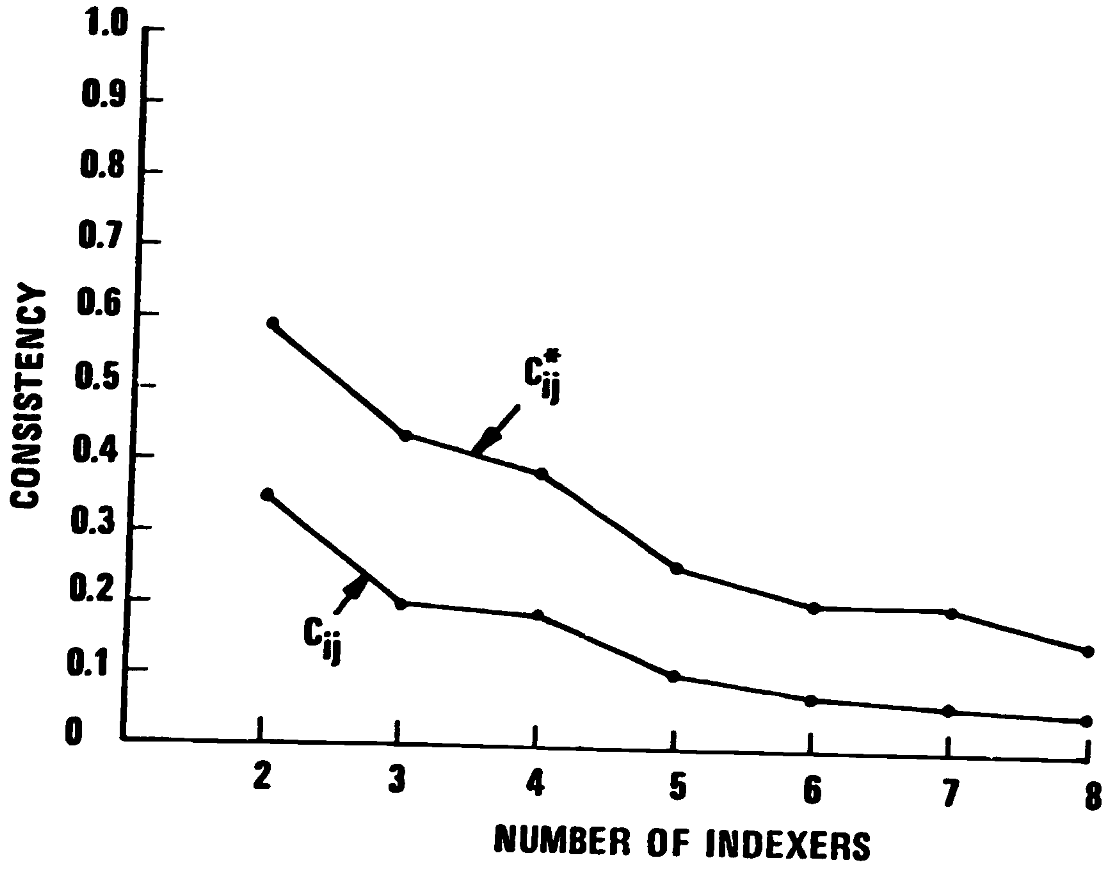
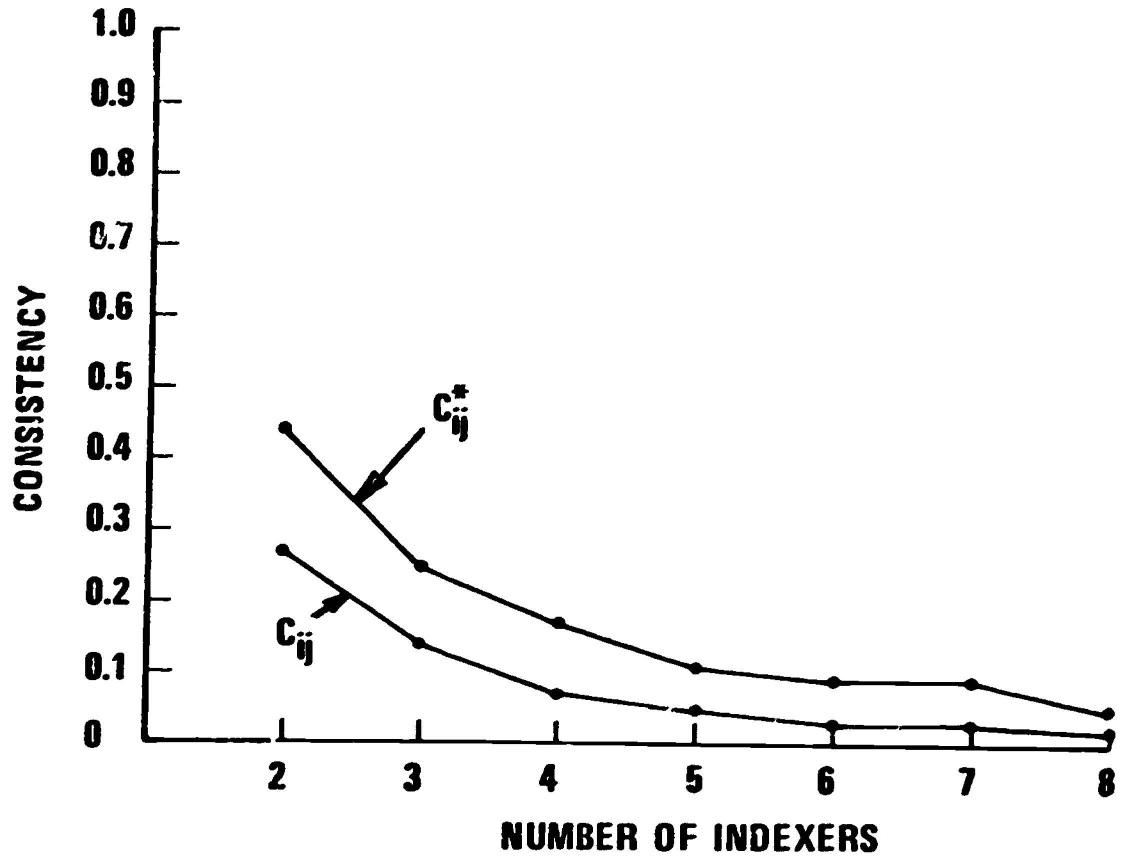


Figure 3. Consistency of Indexing: Values averaged over combinations of indexers and the sample set of documents



Quality Measurements. The indexing quality was calculated, using Eq (4), for each indexer on each document for the professional indexers, the scientists, and the students. The "best" and "worst" indexer in each of the three groups was determined using the average quality for the indexer.

The calculated quality measure for eight professional indexers and eight scientists, based on a sample of 29 documents, and for eight student indexers based on a sample of 16 documents is shown in Tables 3, 4, and 5. The documents were ranked for each group according to the difference in quality of the best indexer and the worst indexer, and the results plotted as shown in figures 4, 5, and 6. It should be noted that the indexer with the best average quality score scored higher in performance quality for most documents individually than the indexer with lowest average quality score. As a matter of fact, there were only two instances where the "poor" indexer of the group performed better than the best in the group by scoring higher on an individual document.*
This is a strong indication that the proposed optimality measure is an adequate tool in evaluating the overall indexer's performance.

The mean and variance of the quality score, and the average number of terms was calculated for all indexers. This data is shown in Table 6. It is interesting to note that the quality measure increases with the number of terms when the number of terms is small. But this increase levels at a given point, and additional terms add little to the quality measure. This seems to suggest an "optimal" number of terms which is, on the average, slightly higher than the average number of terms selected by one indexer.

---

*Two statistical tests were employed, a sign test and the Wilcoxon matched-pairs signed-rank test. Under both tests, the difference in quality between the best and the worst indexer is significant at the .001 level for all three groups of indexers.

Figure 4. Quality Scores of the First Indexer

Group (Professionals) by Document

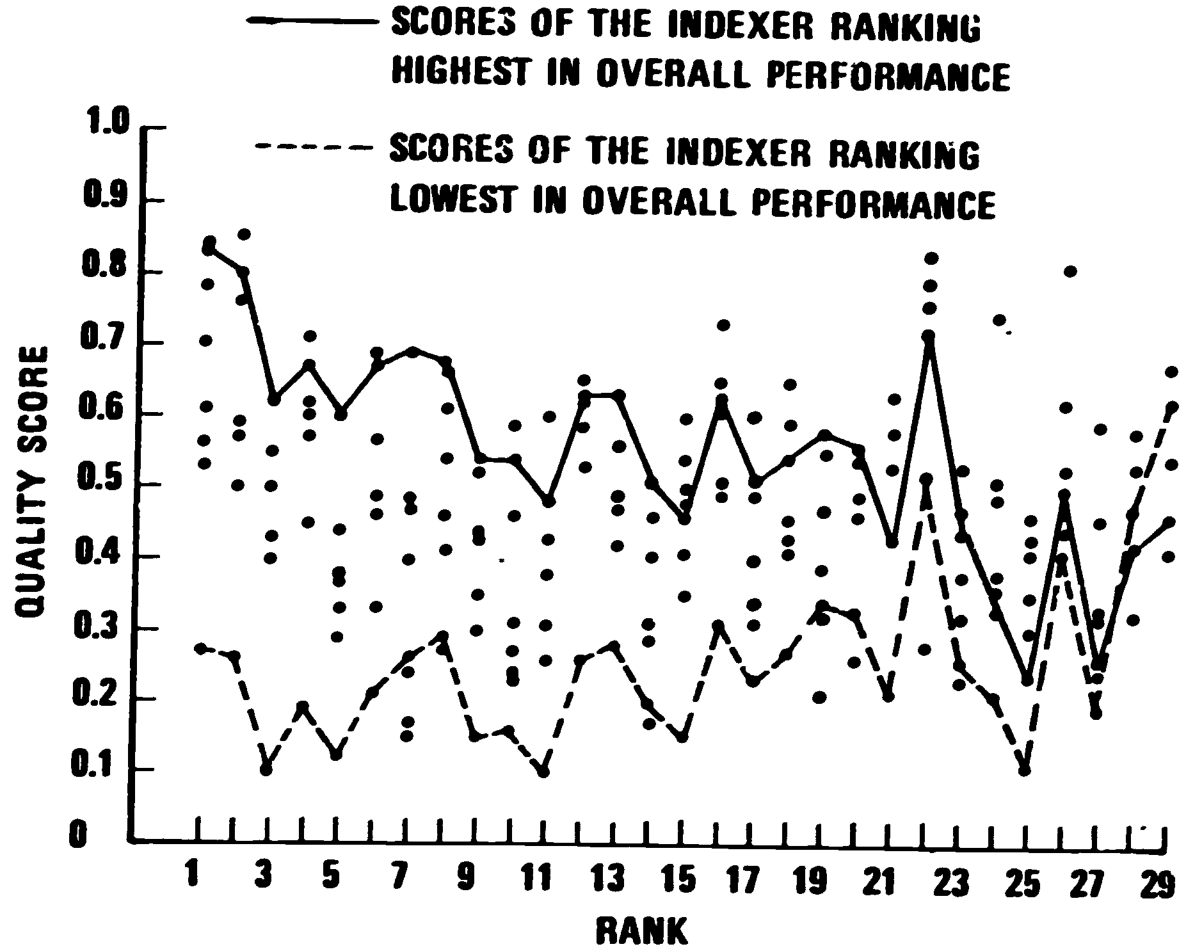


Figure 5. Quality Scores of the Second Indexer Group (Scientists) by Document

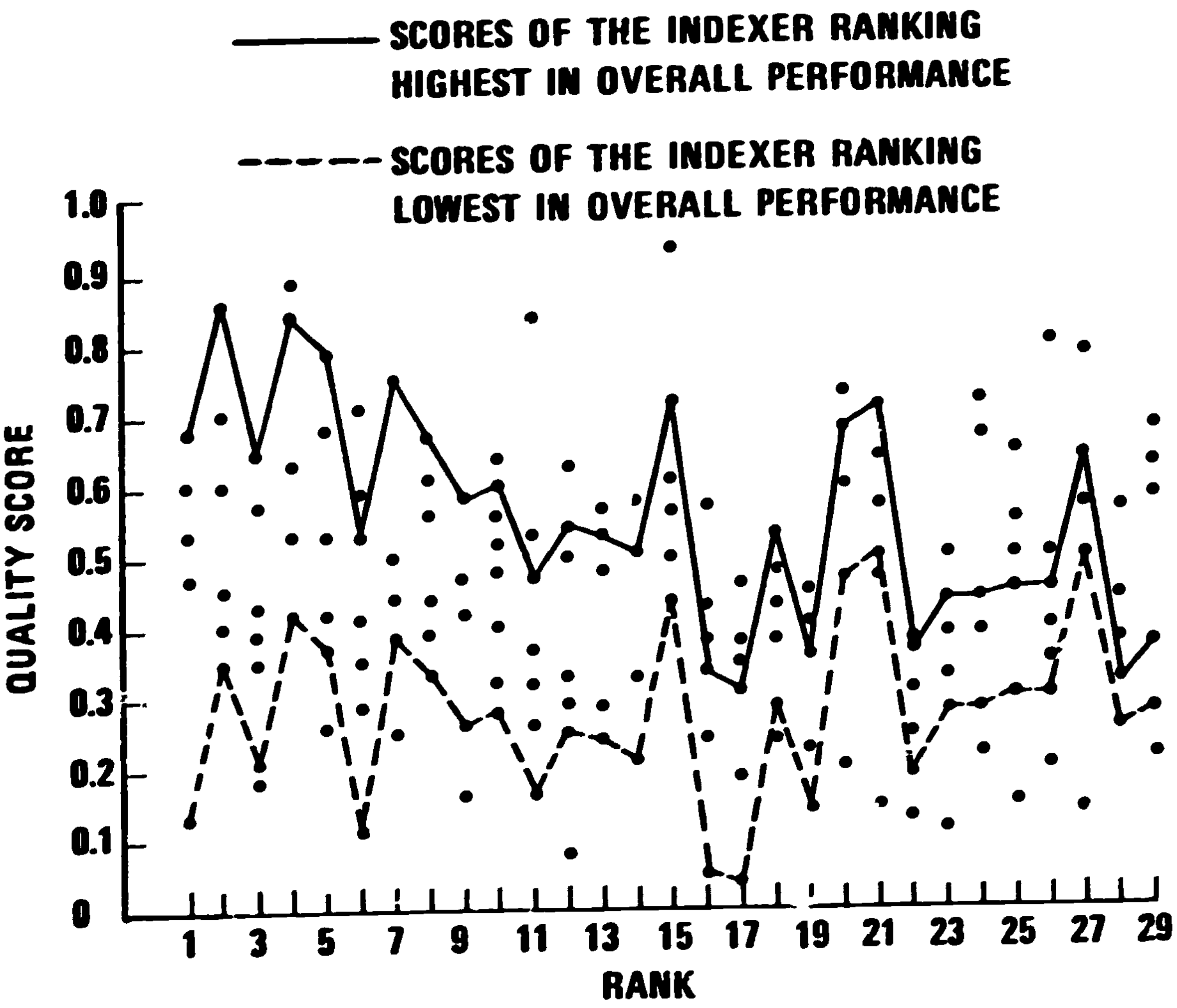


Figure 6. Quality Scores of the Third Indexer Group

(Students) by Document

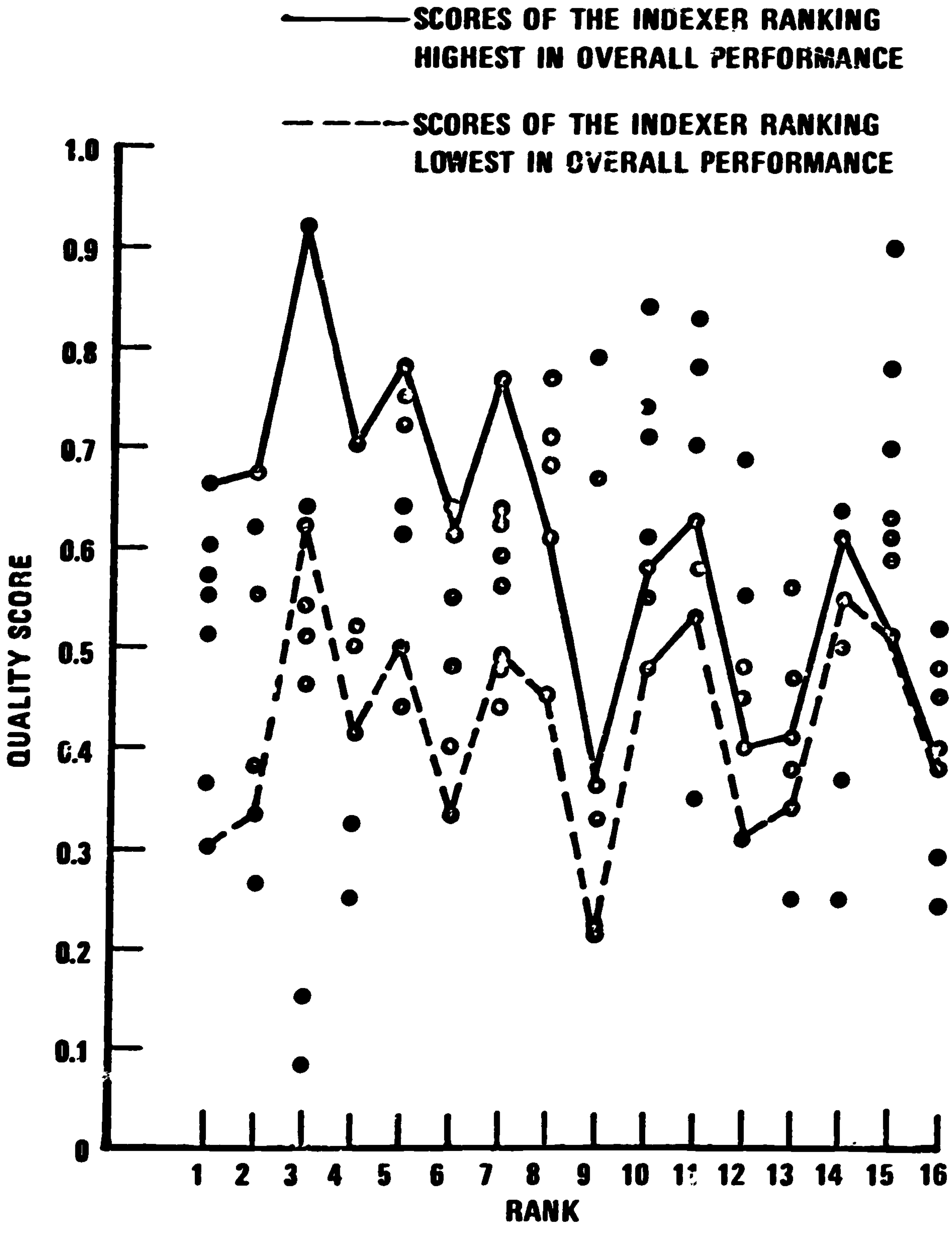


Table 3. Quality Scores by Indexers (Professional) and Documents

		Indexer							
Doc. #	Rank	1	2	3	4	5	6	7	8
6	1	.78	.70	.53	.27	.83	.61	.84	.56
8	2	.59	.85	.50	.26	.80	.50	.57	.76
15	3	.43	.40	.55	.10	.62	.43	.40	.50
7	4	.57	.62	.60	.19	.67	.67	.45	.71
9	5	.38	.33	.44	.12	.60	.37	.29	.44
19	6	.69	.56	.33	.21	.67	.46	.49	.46
2	7	.49	.40	.17	.26	.69	.47	.14	.24
5	8	.61	.27	.41	.29	.68	.46	.66	.54
18	9	.44	.43	.43	.15	.54	.30	.35	.52
4	10	.59	.27	.23	.16	.54	.31	.46	.24
12	11	.38	.31	.60	.10	.48	.31	.26	.43
25	12	.53	.62	.65	.26	.62	.59	.53	.53
30	13	.63	.42	.47	.28	.63	.56	.42	.49
13	14	.31	.29	.40	.20	.51	.31	.17	.46
28	15	.41	.35	.54	.15	.46	.48	.60	.50
1	16	.51	.73	.61	.31	.63	.73	.49	.65
11	17	.34	.51	.60	.23	.51	.31	.49	.40
23	18	.43	.41	.65	.27	.54	.46	.46	.59
14	19	.47	.21	.32	.34	.58	.39	.34	.55
10	20	.46	.54	.26	.33	.56	.49	.33	.49
24	21	.53	.42	.63	.21	.42	.53	.53	.58

Table 3--Continued

		Indexer							
Doc. #	Rank	1	2	3	4	5	6	7	8
21	22	.83	.79	.76	.52	.72	.72	.28	.52
26	23	.23	.32	.47	.26	.44	.38	.53	.47
17	24	.38	.38	.64	.21	.36	.49	.33	.51
29	25	.30	.35	.43	.11	.24	.44	.46	.41
22	26	.62	.44	.81	.41	.50	.53	.47	.47
20	27	.59	.32	.43	.19	.24	.27	.32	.46
16	28	.53	.58	.42	.47	.42	.53	.53	.58
27	29	.67	.54	.64	.62	.46	.41	.54	.62
Average		.52	.44	.50	.26	.53	.47	.44	.51

Table 4. Quality Scores by Indexers (Scientists) and Documents

		Indexer							
Doc. #	Rank	1	2	3	4	5	6	7	8
13	1	.53	.13	.60	.60	.53	.67	.53	.47
22	2	.45	.35	.70	.70	.35	.85	.60	.40
5	3	.57	.21	.18	.43	.39	.64	.35	.21
25	4	.84	.42	.42	.89	.53	.84	.63	.42
24	5	.68	.37	.53	.26	.53	.79	.42	.37
12	6	.35	.12	.41	.59	.29	.53	.71	.35
26	7	.38	.38	.25	.56	.50	.75	.44	.38
20	8	.67	.33	.44	.39	.56	.67	.61	.33
10	9	.42	.26	.26	.47	.26	.58	.16	.47
28	10	.64	.28	.40	.52	.48	.60	.56	.32
23	11	.53	.16	.32	.84	.37	.47	.37	.26
1	12	.63	.25	.29	.50	.63	.54	.08	.33
15	13	.52	.24	.24	.48	.24	.53	.57	.29
27	14	.50	.21	.33	.33	.58	.50	.50	.21
16	15	.61	.44	.44	.94	.50	.72	.56	.44
18	16	.24	.05	.36	.57	.24	.33	.38	.43
2	17	.46	.04	.19	.38	.31	.31	.35	.19
6	18	.48	.29	.43	.38	.48	.53	.24	.38
4	19	.41	.14	.23	.45	.41	.36	.14	.14
14	20	.73	.47	.47	.60	.20	.68	.47	.47
19	21	.64	.50	.50	.57	.57	.71	.14	.50

Table 4--Continued

		Indexer							
Doc. #	Rank	1	2	3	4	5	6	7	8
11	22	.38	.19	.19	.31	.25	.36	.13	.19
9	23	.28	.28	.33	.50	.11	.44	.11	.39
21	24	.67	.28	.44	.72	.22	.44	.39	.28
7	25	.65	.30	.30	.50	.55	.45	.15	.30
29	26	.80	.30	.35	.20	.50	.45	.40	.30
17	27	.64	.50	.50	.79	.57	.64	.14	.50
8	28	.31	.25	.25	.38		.31	.44	.56
30	29	.58	.37	.68	.63	.68	.37	.63	.21
Average		.52	.28	.38	.53	.42	.55	.39	.36

Table 5. Quality Scores by Indexers (Students) and Documents

		Indexer							
Doc. #	Rank	1	2	3	4	5	6	7	8
14	1	.51	.66	.57	.60	.36	.60	.30	.55
9	2	.38	.67	.26	.38	.62	.38	.33	.55
10	3	.08	.92	.64	.15	.54	.51	.62	.46
12	4	.32	.70	.52	.32	.25	.52	.41	.50
7	5	.72	.78	.44	.72	.61	.75	.50	.64
15	6	.55	.61	.40	.55	.48	.64	.33	.55
16	7	.59	.77	.64	.48	.56	.62	.49	.44
6	8	.68	.61	.71	.77	.61	.77	.45	.71
2	9	.22	.36	.33	.67	.33	.79	.21	.33
4	10	.74	.58	.48	.84	.55	.71	.48	.61
11	11	.35	.63	.58	.70	.63	.78	.53	.83
1	12	.55	.40	.48	.55	.45	.69	.31	.48
8	13	.41	.41	.25	.56	.56	.47	.34	.38
13	14	.61	.61	.25	.50	.64	.37	.55	.61
3	15	.61	.51	.90	.70	.63	.59	.51	.78
5	16	.40	.38	.24	.45	.52	.48	.38	.38
Average		.48	.60	.48	.56	.52	.60	.43	.55

Table 6. Average Number of Indexing Terms Assigned, Mean Quality Scores and Variances for Professional Indexers, Scientists, and Student Indexers

	Professional Indexers			Scientists			Students		
		Quality			Quality			Quality	
Rank	No. of Terms	u	6^2	No. of Terms	u	6^2	No. of Terms	u	6^2
1	7.6	.53	.02	3.21	.55	.02	6.3	.60	.02
2	6.7	.52	.03	3.72	.53	.03	6.4	.60	.02
3	5.4	.51	.01	3.38	.52	.02	5.0	.56	.03
4	6.9	.50	.03	2.54	.42	.02	4.4	.55	.02
5	5.9	.47	.03	2.24	.39	.03	3.2	.43	.01
6	4.4	.44	.04	1.76	.38	.02	4.9	.48	.03
7	5.0	.44	.03	1.48	.36	.02	4.5	.48	.03
8	2.5	.26	.01	1.0	.28	.01	3.2	.43	.01
AVG	5.6			2.4			5.0		

<u>Measurement of Indexing Exhaustivity</u>. The question to be discussed next is how much an index can be improved by "group" indexing, i.e. by having one and the same document indexed by several indexers and the combining the results into a single index for that document. In other words, the problem under investigation was how much more information is contained in the joint product of two indexers as compared to one indexer, or in the joint product of three indexers as compared to two, etc.

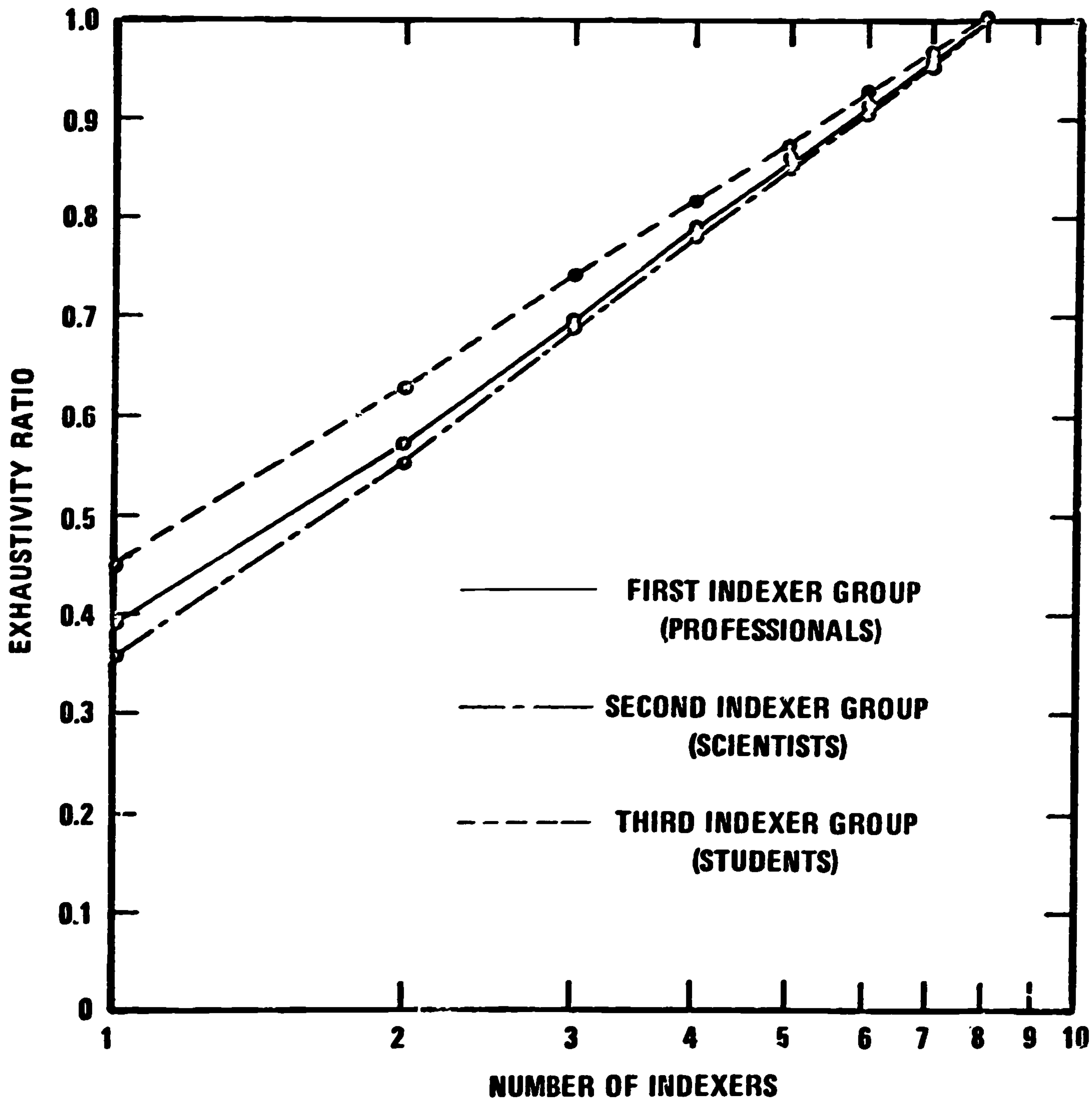
One approach to answering this question is to consider the number of new terms contributed by another indexer. Calculations were made as to the average number of terms selected by one indexer, two indexers, and so on, out of the total set of terms selected by all eight indexers by considering all combinations of indexers two at a time, three at a time, etc. These values were then expressed in percentages of the total number of terms resulting from all eight indexers. This data is shown in Table 7. Although the number of terms were quite different, the percentages obtained were practically identical for the scientists and professional indexers, and not greatly different for the students. In all cases, approximately half the terms obtained by eight indexers were obtained by two indexers, and 80% of the terms were obtained by five.

The calculations were repeated using the proposed measure of exhaustivity Eq (4), which reflects the varying degrees of significance of information contents of assigned terms. Average exhaustivity ratios were calculated for groups of 1, 2, 3, ..., to 8 indexers and the results are shown in Table 7, columns 3, 5, and 7. The plot of these exhaustivity ratios on a logrithmic scale is shown in Figure 7.

Table 7. Percentages of the Total Number of Terms and Exhaustivity Levels as a Function of the Number of Indexers in the Group

	Professional Indexers		Scientists		Students	
Number of Indexers	Percentage of Terms	100.γ	Percentage of Terms	100.γ	Percentage of Terms	100.γ
1	29%	39%	28%	36%	33%	45%
2	46%	57%	46%	55%	52%	63%
3	59%	70%	60%	69%	64%	74%
4	70%	79%	71%	78%	72%	82%
5	79%	86%	80%	86%	81%	88%
6	87%	91%	88%	92%	88%	93%
7	94%	96%	94%	96%	94%	97%
8	100%	100%	100%	100%	100%	100%

Figure 7. Plot of the Exhaustivity Ratios γ Versus the Number of Indexers in the Group (on a log scale) for a Group of 8 Indexers



The interesting result of this experiment is that the exhaustivity ratios thus obtained, when plotted on a logarithmic scale with respect to the number of indexers, lie on a straight line for each group of indexers considered. This suggests a general model of indexing exhaustivity

$$\gamma = \gamma_1 + m \log n \qquad \text{and} \qquad N << T \qquad (8)$$

where γ is the exhaustivity measure, γ_1 is the "average exhaustivity" of one indexer (intercept of the line with the n-axis), n is the number of indexers, $m = \tan \phi$ is the tangent of the angle at which the line intersects the n axis, N is the total number of indexers in the group and T is the total number of different indexing terms used by the group.

To test this model, another calculation was performed independently for a group of 20 indexers. The results are shown in the plot of Fig. 8 and can be seen to agree well with the previous findings. However, as N approaches T, the dependence of the exhaustivity coefficient γ on the number of indexers n might be expected to deviate more and more from the expression Eq (8) because — and this is intuitively clear — the distribution law underlying this process would not hold unless N << T.

Since

$$\gamma(n+1) = \gamma_1 + m \log (n+1)$$

therefore the difference or the rank increment of γ

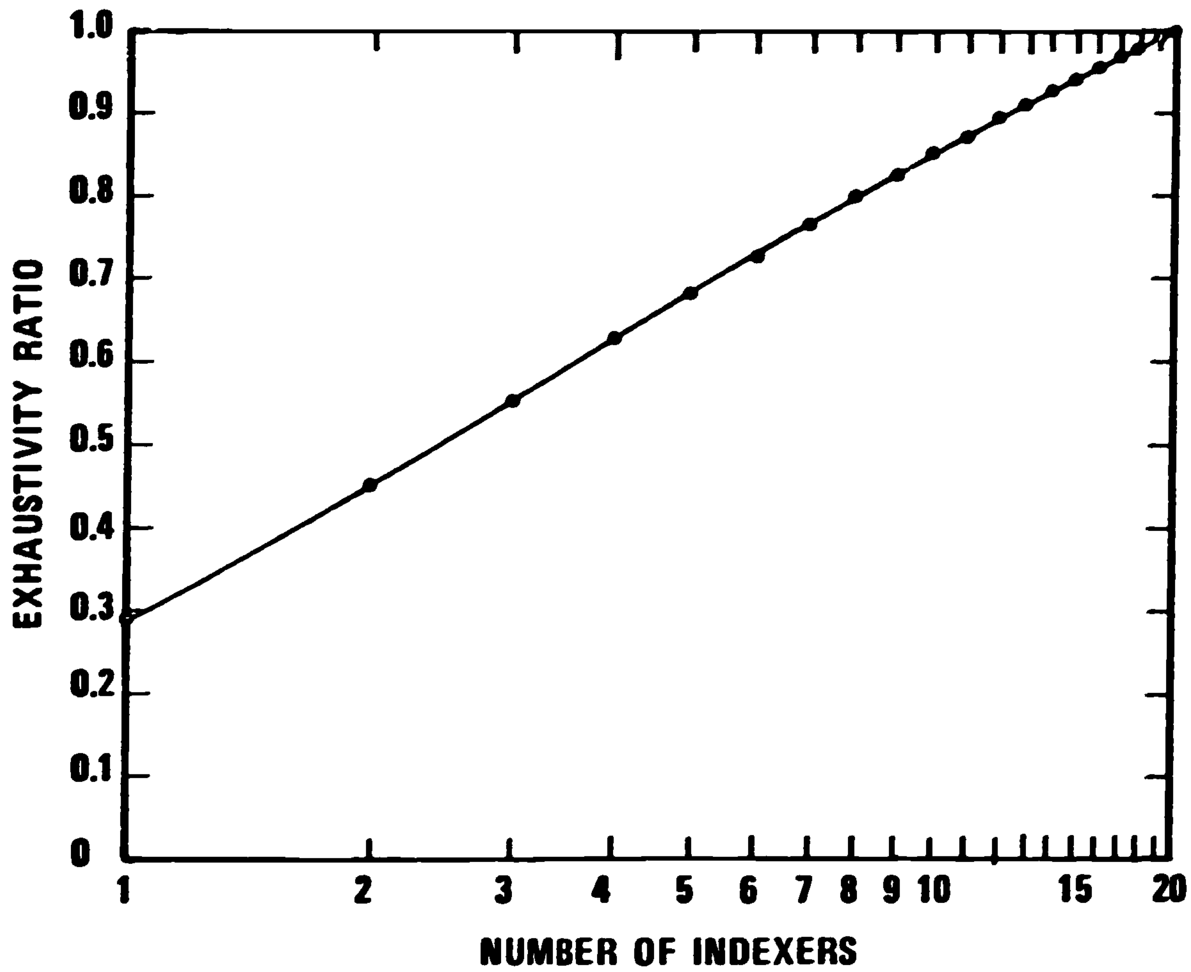
$$\Delta\gamma_n = (n+1) - \gamma(n) = m \log (n+1) - m \log n$$

can be approximately equated to

$$\Delta\gamma_n = m \log (1 + \frac{1}{n}) \cong \frac{m}{\ln 10} \cdot \frac{1}{n} = \hat{m} \frac{1}{n}$$

or

Figure 8. Plot of the Exhaustivity Ratios γ Versus the Number of Indexers in the Group (on a log scale) for a Group of 20 Indexers



$$\Delta\gamma_n \cdot n = m = \text{const.}$$

Thus Eq (8) or its equivalent form Eq (9) are familiar expressions obtained in other contexts: they can be interpreted — with appropriate changes of the names of variables — either as Bradford's law of information scattering (14), as the law of scientific productivity developed by Lotka (15), or as Zipf's law of vocabulary distribution (16). In Bradford's law, the expression of the form of Eq (8) relates the cumulative total number of articles to the number of journals in which they were published. Lotka obtained an equivalent functional relationship (although in more general form) between the number of scientists and the number of scientific contributions they make. In the case of Zipf, an expression of the type of Eq (9) was obtained relating the word length with the number of word occurrences. We have shown that the same model applies also to express the relation between the level of exhausitivity of information representation or indexing and the number of decision makers (or indexers) whose individual judgments are combined to reproduce the information contents of the object which is jointly observed (specifically, to produce a cumulative index). In other words, this could be considered as the law expressing the dependence of the "completeness" of knowledge on the size of the community which might be expected to contribute to this knowledge.

The fact that the same model is applicable to the process of information scattering (Bradford's Law), growth of scientific productivity (Lotka's Law), vocabulary distribution (Zipf's Law), and, as we have now shown, to the completeness or exhaustivity of information representation as reflected in group indexing, implies that the same basic law of distribution of information flow should underlie all these processes. This topic will be discussed in greater detail in a different paper.

Conclusions. Measures of indexing consistency should reflect not only the formal agreement of indexers on a number of terms, but also the significance of terms on which the indexers agree or disagree. This can be achieved if the sets of indexing terms are considered not well-defined, but fuzzy sets with respect to the significance judgment. A procedure for designing such sets has been proposed and it has been shown that the proposed approach can be extended to define the quality and exhaustivity of indexing. The experimental results on indexing exhaustivity based on this approach warrant the conclusion that the increase of information as a result of group indexing obeys the same law as that of the scattering of information (Bradford), of scientific productivity (Lotka), and vocabulary distribution (Zipf).

REFERENCES

1. Hooper, R. S., Indexer Consistency Tests - Origin, Measurements, Results and Utilizatio[illegible]. IBM Washington Systems Center, Bethesda, Maryland, P[illegible] [illegible]ted at the 1965 Congress International Federation f[illegible] - [illegible]umentation 10-15 October, 1965.

2. Jacoby, J., Methodology f[illegible] [illegible]ndexer Reliability Texts, (RADC-IN-62-1), Documentation, I[illegible]., Bethesda, Maryland, March, 1962.

3. Slamecka, V., and J. Jacoby, Effect of Indexing Aids on the Reliability of Indexers, (RADC-TDR-63-116), Documentation, Inc., Bethesda, Maryland, June, 1963.

4. Jacoby, J. and V. Slamecka, Indexer Consistency under Minimal Conditions, (RADC-TDR-62-426), Documentation, Inc., Bethesda, Maryland, November, 1962.

5. Painter, A. F., An Analysis of Duplication and Consistency of Subject Indexing Involved in Report Handling at the Office of Technical Services, U.S. Department of Commerce, Office of Technical Services, Washington, D.C., March, 1963.

6. Rodgers, D. J., A Study of Intra-Indexer Consistency, General Electric Company, Washington, D.C., September, 1961.

7. Korotkin, A. L., and L. H. Oliver, The Effect of Subject Matter Familiarity and the Use of an Indexing Aid upon Inter-Indexer Consistency, General Electric Company; Bethesda, Maryland, February, 1964.

8. Hooper, R. S., A Facet Analysis System, Automation and Scientific Communication, Short Papers, Annual Meeting of the American Documentation Institute, Vol. 2, p. 253, October, 1963.

9. Stevens, M. E., Automatic Indexing: A State-of-the-Art Report, Washington, D.C., National Bureau of Standards, Monograph 91, p. 160, March, 1965.

10. Bryant, E. C., D. W. King, and P. J. Terragno, Some Technical Notes on Coding Errors, Denver, Colorado, Westat Research Analysts, 1963.

11. Schultz, Claire K., Wallace L. Schultz, and Richard H. Orr, Comparative Indexing: Terms Supplied by Biomedical Authors and by Document Titles, American Documentation, 16(No. 4): 299-312 (1965).

12. Zadeh, L. A., Fuzzy Sets, Information and Control 8: 338-352 (1965).

13. Brown, James E., Efficient Reading, D. C. Heath and Company, Boston (1962).

14. Bradford, S. C., Documentation, Public Affairs Press, Washington D.C., 1950, pp. 148-154.

15. Lotka, Alfred J., The Frequency Distribution of Scientific Productivity, Journal of the Washington Academy of Sciences, 16 (No. 12) 317-323, (1926).

16. Zipf, G. K., Human Behavior and the Principle of Least Effort, Addison-Wesley Press, Inc., Cambridge, Mass., 1949, pp.24-25.

APPENDIX

Operations on Fuzzy Sets. The purpose of this appendix is to expand the concept of fuzzy sets discussed in the text. The definitions are from Zadeh (12).

Let A, B and C be fuzzy sets in a set X of objects x with membership functions $f_A(x)$, $f_B(x)$ and $f_C(x)$ respectively. Then the following definitions are made:

Equality: The fuzzy sets A and B are said to be equal, written $A = B$, if $f_A(x) = f_B(x)$ for all x in X.

Subsets: The fuzzy set A is said to be a subset of the fuzzy set B if $f_A(x) \leq f_B(x)$ for all x in X.

Union: The union of two fuzzy sets A and B, written $A \cup B$, is a fuzzy set C where $f_C(x) = \mathrm{Max}\ [f_A(x), f_B(x)]$, $x \in A$ which may be abbreviated $f_C = f_A \vee f_B$. The union of n fuzzy sets $T_1, T_2, \ldots T_n$ is the fuzzy set C where $f_C = f_{T_1} \vee f_{T_2} \vee \ldots \vee f_{T_n} = \underset{i}{\vee} f_{T_i}$

Intersection: The intersection of two fuzzy sets A and B, written $A \cap B$, is a fuzzy set C where $f_C(x) = \mathrm{Min}\ [f_A(x), f_B(x)]$, $x \in A$ which may be abbreviated $f_C = f_A \wedge f_B$. The intersection of n fuzzy sets $T_1, T_2, \ldots, T_n$ is the fuzzy set C where $f_C = f_{T_1} \wedge f_{T_2} \wedge \ldots \wedge f_{T_n} = \underset{i}{\wedge} f_{T_i}$

It can be shown that fuzzy sets have the following properties:

$A \cup B = B \cup A$	$A \cap B = B \cap A$
$(A \cup B) \cup C = A \cup (B \cup C)$	$(A \cap B) \cap C = A \cap (B \cap C)$
$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$	$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
$A \cup A = A$	$A \cap A = A$